

Combining Concepts and Language Models for Information Access

Edgar Meij

University of Amsterdam, The Netherlands

edgar.meij@uva.nl

Abstract

Since the middle of last century, information retrieval has gained an increasing interest. Since its inception, much research has been devoted to finding optimal ways of representing both documents and queries, as well as improving ways of matching one with the other. In cases where document annotations or explicit semantics are available, matching algorithms can be informed using the concept languages in which such semantics are usually defined. These algorithms are able to match queries and documents based on textual and semantic evidence.

Recent advances have enabled the use of rich query representations in the form of query language models. This, in turn, allows us to account for the language associated with concepts within the retrieval model, in a principled and transparent manner. Developments in the semantic web community, such as the Linked Open Data cloud, have enabled the association of texts with concepts on a large scale. Taken together, these developments facilitate a move beyond manually assigned concepts in domain-specific contexts into the general domain.

This thesis investigates how one can improve information access by employing the actual use of concepts as measured by the language that people use when they discuss them. The main contribution is a set of models and methods that enable users to retrieve and access information on a conceptual level. Through extensive evaluations, a systematic exploration and thorough analysis of the experimental results of the proposed models is performed. Our empirical results show that a combination of top-down conceptual information and bottom-up statistical information obtains optimal performance on a variety of tasks and test collections.

The dissertation is available online at <http://phdthes.is/>.
